



ACADGILD

SESSION 2: INTRODUCTION to working with R

Assignment 2

Problem Statement

1. Read multiple json files into a working directory for further converting into a dataset.

I have files text1, text2, text3 in the directory json.

2. Parse the following JSON into a data frame

```
js<-'{  
  "name": null, "release_date_local": null, "title": "3 (2011)",  
  "opening_weekend_take": 1234, "year": 2011,  
  "release_date_wide": "2011-09-16", "gross": 59954  
}'
```

3. Write a script for variable binning using R.

Answers :

```
# Acadgild-Data-Analytics-Session-2_Assignment-2.2
```

SOLUTION :

1. The sample json files text1.json, text2.json and text3.json are present in the folder “json” in

F:\ACADGILD - Online Course\\1. DATA SETS

Jsonlite and dplyr packages are installed and then following commands are executed using R-studio:

```
# Reading multiple files using for loop and convert into a dataset
```

```
library(jsonlite)
```

```

library(dplyr)

ls <- list("F:\\ACADGILD - Online Course\\1. DATA SETS\\text1.json",
          "F:\\ACADGILD - Online Course\\1. DATA SETS\\text2.json",
          "F:\\ACADGILD - Online Course\\1. DATA SETS\\text3.json")

for (i in ls){
  a[i] <- read_json(i, simplifyVector = TRUE)

  z[i] <- data.frame( i,row.names = NULL, check.rows = FALSE,
                     check.names = TRUE, fix.empty.names = TRUE,
                     stringsAsFactors = default.stringsAsFactors())

  z[i] <- cbind(z[i],a[i])
}

View(a)
View(z)

```

Hence multiple json files are read into the working directory and are then converted into datasets.

The current working directory may be obtained by using getwd()

```

# Reading multiple files one by one and convert into a dataset

library(jsonlite)
library(dplyr)

y1<-read_json("F:\\ACADGILD - Online Course\\1. DATA SETS\\text1.json")
View(y1)

result1<- as.data.frame(do.call("rbind",y1))
result1

y2<-read_json("F:\\ACADGILD - Online Course\\1. DATA SETS\\text2.json")

```

View(y2)

```
result2<- as.data.frame(do.call("rbind",y2))
```

result2

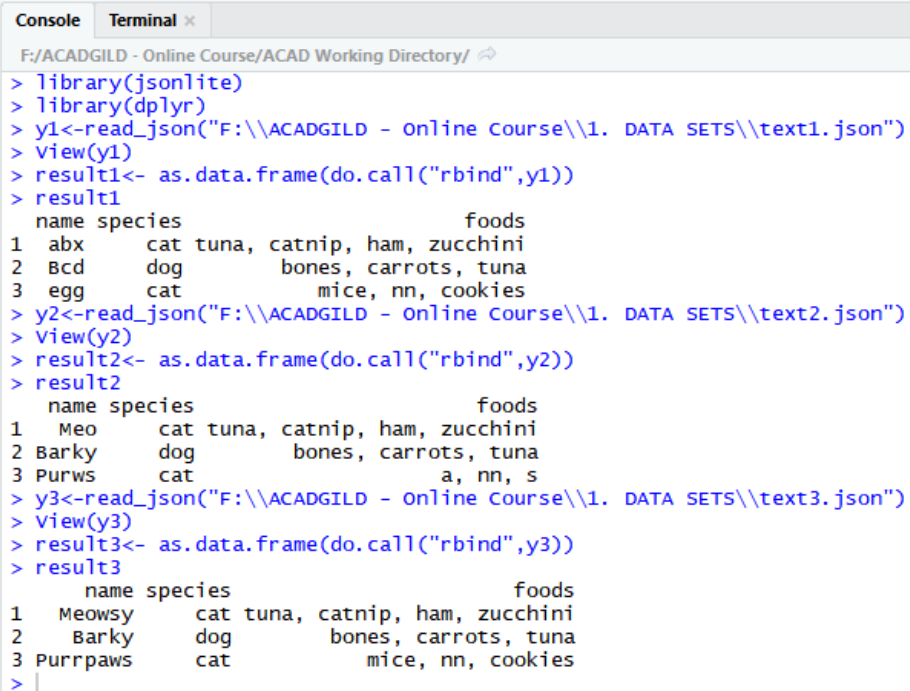
```
y3<-read_json("F:\\ACADGILD - Online Course\\1. DATA SETS\\text3.json")
```

View(y3)

```
result3<- as.data.frame(do.call("rbind",y3))
```

result3

The output is shown as below :



```
Console Terminal x
F:/ACADGILD - Online Course/ACAD Working Directory/
> library(jsonlite)
> library(dplyr)
> y1<-read_json("F:\\ACADGILD - Online Course\\1. DATA SETS\\text1.json")
> view(y1)
> result1<- as.data.frame(do.call("rbind",y1))
> result1
  name species              foods
1  abx      cat tuna, catnip, ham, zucchini
2  Bcd      dog      bones, carrots, tuna
3  egg      cat      mice, nn, cookies
> y2<-read_json("F:\\ACADGILD - Online Course\\1. DATA SETS\\text2.json")
> view(y2)
> result2<- as.data.frame(do.call("rbind",y2))
> result2
  name species              foods
1  Meo      cat tuna, catnip, ham, zucchini
2  Barky    dog      bones, carrots, tuna
3  Purws    cat      a, nn, s
> y3<-read_json("F:\\ACADGILD - Online Course\\1. DATA SETS\\text3.json")
> view(y3)
> result3<- as.data.frame(do.call("rbind",y3))
> result3
  name species              foods
1  Meowsy    cat tuna, catnip, ham, zucchini
2  Barky    dog      bones, carrots, tuna
3  Purrpaws  cat      mice, nn, cookies
> |
```

F:/ACADGILD - Online Course/ACAD Working Directory - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Untitled1* x Untitled3* x Untitled4* x y3 x y2 x y1 x

Show Attributes

Name	Type	Value
y1	list [3]	List of length 3
[[1]]	list [3]	List of length 3
[[2]]	list [3]	List of length 3
[[3]]	list [3]	List of length 3

(No selection)

Console

F:/ACADGILD - Online Course/ACAD Working Directory - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

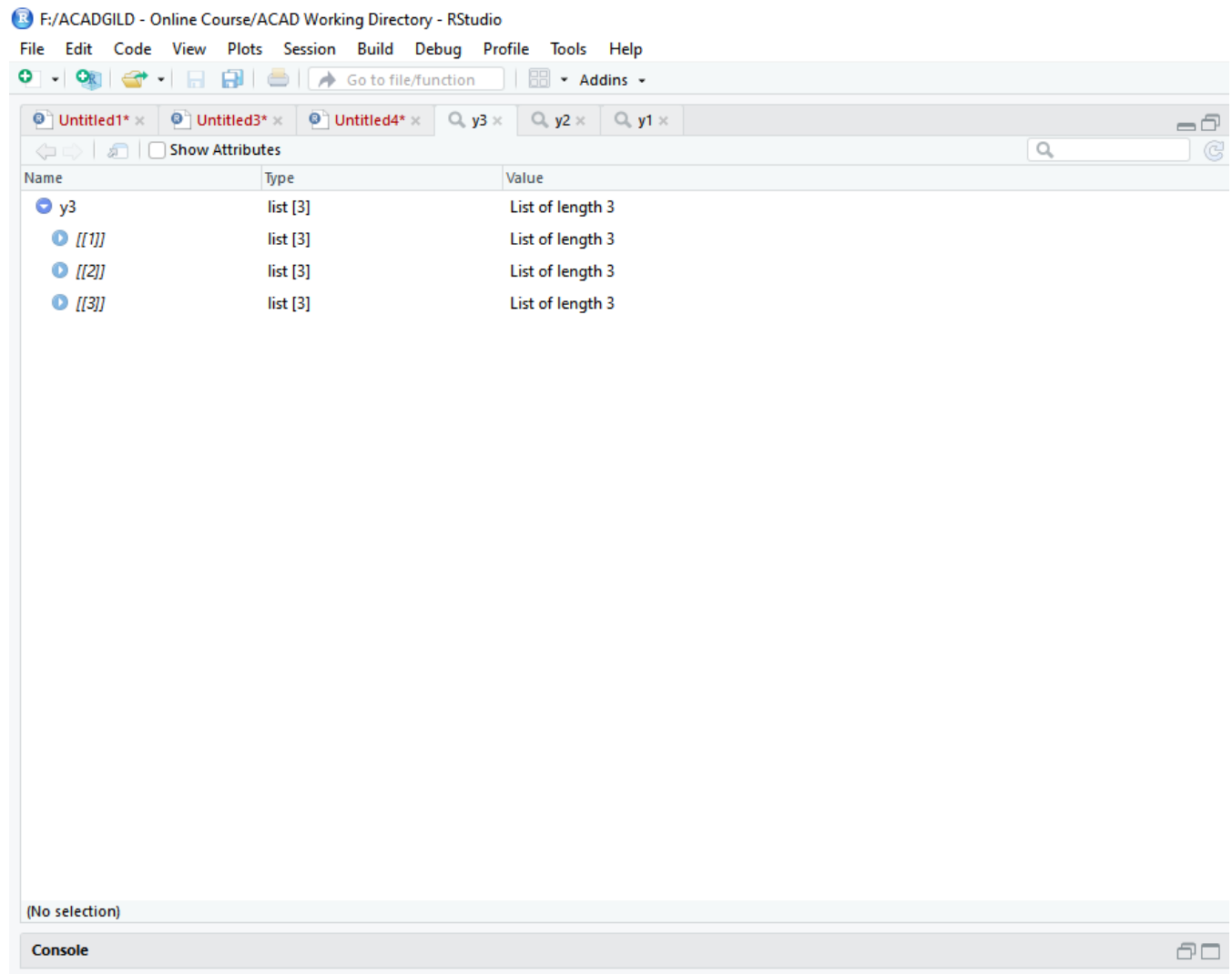
Untitled1* x Untitled3* x Untitled4* x y3 x y2 x y1 x

Show Attributes

Name	Type	Value
y2	list [3]	List of length 3
[[1]]	list [3]	List of length 3
[[2]]	list [3]	List of length 3
[[3]]	list [3]	List of length 3

(No selection)

Console



2.

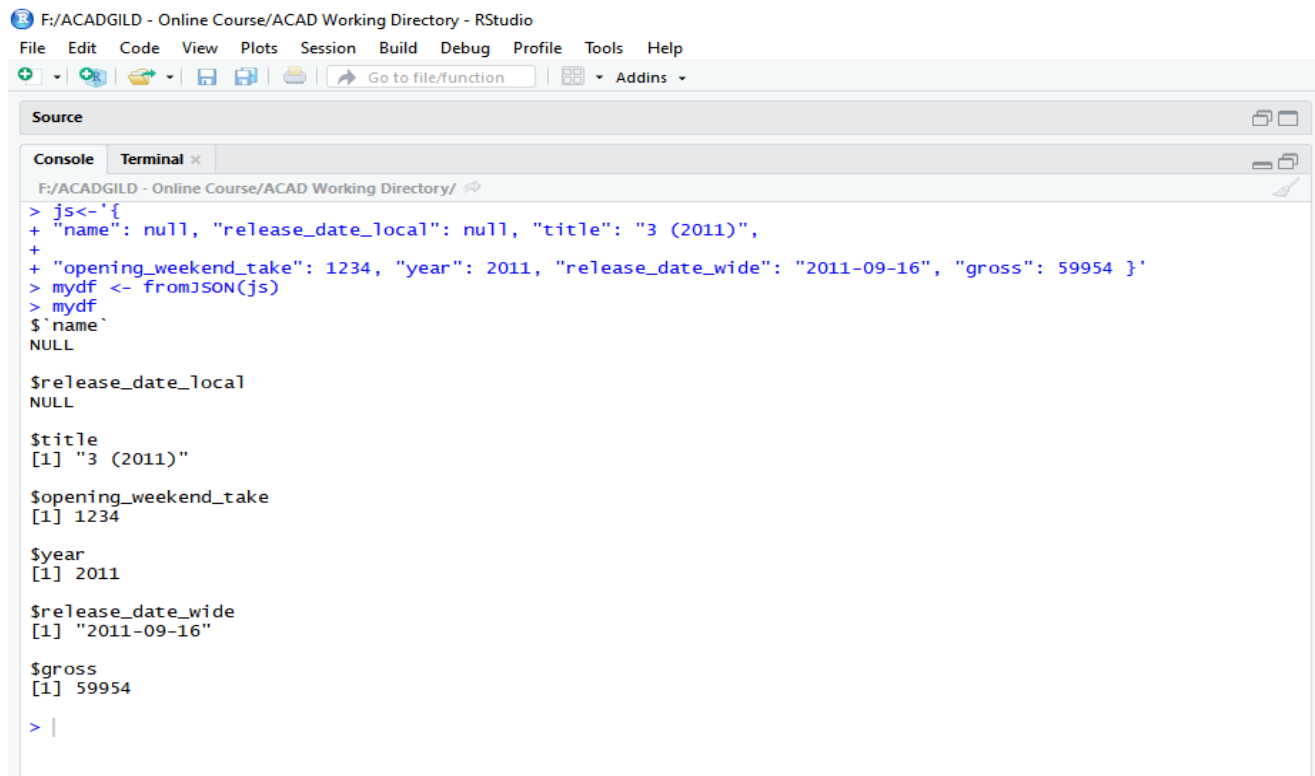
```
js<-'{  
  "name": null, "release_date_local": null, "title": "3 (2011)",  
  "opening_weekend_take": 1234, "year": 2011,  
  "release_date_wide": "2011-09-16", "gross": 59954  
'
```

```
mydf <- fromJSON(js)
```

```
mydf
```

Here the given Jason is stored in variable named js. fromJSON() function is used for the parsing the data into dataframe. The resultant data frame is stored in mydf.

The output is given below:



The screenshot shows the RStudio interface with the console pane active. The title bar indicates the working directory is 'F:/ACADGILD - Online Course/ACAD Working Directory - RStudio'. The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar shows various icons for file operations and a 'Go to file/function' search bar. The console pane displays the following R code and its output:

```
> js<-'{
+ "name": null, "release_date_local": null, "title": "3 (2011)",
+ "opening_weekend_take": 1234, "year": 2011, "release_date_wide": "2011-09-16", "gross": 59954 }'
> mydf <- fromJSON(js)
> mydf
$name
NULL

$release_date_local
NULL

$title
[1] "3 (2011)"

$opening_weekend_take
[1] 1234

$year
[1] 2011

$release_date_wide
[1] "2011-09-16"

$gross
[1] 59954

> |
```

3.

Binning is the process of transforming numerical variables into categorical counterparts.

VARIABLE BINNING USING cut() function

```
v <- 1:400
print(v)
v <- seq(1:100)
print(v)
tapply(v, cut(v, 60))
```

The output is given below:

Source

Console

Terminal x

F:/ACADGILD - Online Course/ACAD Working Directory/

```

> v <-1:400
> print(v)
 [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
[26] 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
[51] 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75
[76] 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
[101] 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125
[126] 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
[151] 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175
[176] 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
[201] 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225
[226] 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250
[251] 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275
[276] 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
[301] 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325
[326] 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350
[351] 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375
[376] 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
> v<-seq(1:100)
> print(v)
 [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
[26] 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
[51] 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75
[76] 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
> tapply(v,cut(v,60))
 [1] 1 1 2 2 3 4 4 5 5 6 7 7 8 8 9 10 10 11 11 12 13 13 14 14 15 16 16 17 17 18 19 19 20 20
[35] 21 22 22 23 24 24 25 25 26 27 27 28 28 29 30 30 31 31 32 33 33 34 34 35 36 36 37 37 38 39 39 40 40 41
[69] 42 42 43 44 44 45 45 46 47 47 48 48 49 50 50 51 51 52 53 53 54 54 55 56 56 57 57 58 59 59 60 60
>

```